Raghavendra R, Rao A, Hemantha Kumar G. Multimodal biometric score fusion using Gaussian mixture model and Monte Carlo method. JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY 25(4): 771–782 July 2010. DOI 10.1007/s11390-010-1060-0

Multimodal Biometric Score Fusion Using Gaussian Mixture Model and Monte Carlo Method

R Raghavendra[1], Rao Ashok[2], and G Hemantha Kumar[1]

[1] *Department of Studies in Computer Science, University of Mysore, Mysore-570 006, India*

[2] *Department of E& C, Channabasaveshwara Institute of Technology, Gubbi-Tumkur-572 216, India*

E-mail: raghu07.mys@gmail.com; ashokrao.mts@gmail.com; hemanthakumar@compsci.uni-mysore.ac.in

Received April 26, 2009; revised February 3, 2010.

Abstract Multimodal biometric fusion is gaining more attention among researchers in recent days. As multimodal biometric system consolidates the information from multiple biometric sources, the effective fusion of information obtained at score level is a challenging task. In this paper, we propose a framework for optimal fusion of match scores based on Gaussian Mixture Model (GMM) and Monte Carlo sampling based hypothesis testing. The proposed fusion approach has the ability to handle: 1) small size of match scores as is more commonly encountered in biometric fusion, and 2) arbitrary distribution of match scores which is more pronounced when discrete scores and multimodal features are present. The proposed fusion scheme is compared with well established schemes such as Likelihood Ratio (LR) method and weighted SUM rule. Extensive experiments carried out on five different multimodal biometric databases indicate that the proposed fusion scheme achieves higher performance as compared with other contemporary state of art fusion techniques.

Keywords multimodal biometric system, match score level fusion, Gaussian mixture model, Monte Carlo method.

1 Introduction

Identification of person with high accuracy is becoming critical in a number of security issues in our society. Person verification based on biometric features has attracted more attention in designing security system. Most of the biometric systems that are presently in use typically have a single biometric trait to establish identity (unimodal biometric systems). Unimodal biometric system suffers from serious drawbacks such as noise in sensed data, lack of universality, and susceptible to spoof attacks[1]. Some of the limitations of Unimodal biometrics are alleviated by combining evidence from more than one source of biometric information. A system that combine more than one biometric trait is termed as multimodal biometric system. This improves the matching accuracy of a system while increasing population coverage and significantly reducing spoof attacks. This also implies complicated modeling, intensive computation and in most cases use of sophisticated algorithms[1-2]. The heart of multimodal biometric system relies on fusing the information from multiple biometric traits in order to achieve better recognition performance. Fusion can be performed at four different levels of information such as sensor level, feature level, match score level, and decision level. In the sensor level fusion, the raw data from different sensors are combined. Sensor level fusion is applicable only if the multiple sources represent samples of the same biometric trait obtained either using a single sensor or different compatible sensor[1]. In feature level fusion, features from different modalities are combined to form a single feature. When feature sets are homogeneous, a single resulting feature vector can be calculated as a weighted average of the individual feature vectors. When features are non-homogeneous, we can concatenate them to form a single vector. In case of score level fusion, the match score given by more than one biometric system are combined to make the decision of accept/reject. Match score level fusion is generally preferred as all commercially available biometric devices may not provide access to the feature information at all levels and also that it is easy to fuse at this level. In decision level fusion, the decision given by the individual biometric system is combined to make the final decision. The decision level fusion is also called as abstract level fusion as we combine the binary information provided by individual biometric system.

There exist a large number of techniques for fusing the scores obtained from different biometric trait.

Regular Paper

Kittler *et al.*[3], presents a theoretical framework for fusion approaches such as sum rule, product rule, median rule, min rule, max rule. In their work[3], experimental results combining the opinion from three different experts in which two are face experts (frontal and profile) and one text dependent speech expert indicates the supreme performance of sum rule. Jain *et al.*[4], proposed a multimodal biometric system based on speech, face and fingerprint. Their match score level fusion is carried out using likelihood ratio that was obtained using Neyman-Pearson rule. Snelick *et al.*[5] compared combination of different normalization and fusion schemes for match score level fusion of face and fingerprint biometric. Ross *et al.*[6] fuse face, fingerprint and hand geometry scores for person verification. Their results based on experiments indicate that the sum rule performs better than the decision tree and linear discriminant classifier. Chu *et al.*[7] present a face and palmprint based multimodal biometric identification method. Their system uses ordinal features of face and palmprint. Here match score level fusion based on the sum rule, product rule, min rule, max rule and Fishers Discriminate Analysis (FLD) is carried out. The outcome from their experiments indicate that the best results are obtained using LDA. Dass *et al.*[8] combine match scores from multiple matchers based on generalized density estimation from the genuine and imposter scores. Aguiler *et al.*[9] combine match scores of fingerprint and online signature using Support Vector Machine (SVM) classifier and better results are reported as compared with individual biometrics. Nandakumar *et al.*[2] proposed a density based score level fusion using likelihood ratio. In their work, Gaussian Mixture Model is used to estimate the likelihood ratio for fusion of scores. In addition, many interesting schemes have been proposed with the best among them being Likelihood Ratio[2] and Weighted SUM Rule[1].

All of the above suffer from two major limitations: 1) the performance while being good is not yet excellent, and 2) not robust enough to be used in many noisy (realistic) situations. In this paper, we do attempt and obtain a significantly improved performance by using novel approach that involves Monte Carlo sampling method. The robustness of performance with respect to noise is outside the scope of this work.

In this paper, we employ match score level fusion. Combining the match scores is a challenging task as scores from different biometric traits may follow unknown distribution and scores from different matcher can provide either distance or similarity measures. Another major problem in score level fusion is that, we do not have access to large amount of scores (especially genuine match scores) to reliably estimate the genuine and imposter scores distribution. One way of addressing this problem is by sampling the distribution of scores obtained from different biometric traits. In this paper, we propose a novel method for fusing the match scores from different biometrics based on GMM and Monte Carlo sampling method. We use GMM to estimate the statistics of match scores which in turn will be used by Monte Carlo method to do sampling. Hitherto the best results[2], which are considered as state-of-the-art, have been obtained using a very complicated and computationally intensive process of likelihood estimation. This is followed by hypothesis testing. We have avoided the complication of estimation of Likelihood Ratio by directly using a more efficient and simpler sampling process. Computation of this is significantly less intensive than likelihood estimation. Finally, we use statistical hypothesis testing on sampled scores to decide whether user is genuine or imposter. The proposed method is validated on five different multimodal biometric databases with varying the number of matchers (biometric trait) and also size.

The rest of the paper is organized as follows. Section 2 describes the proposed method; Section 3 describes the experimental setup; Section 4 describes the results and discussion; Section 5 draws the conclusion. Appendix elaborates the various issues related to experiments done like:

1) Need for estimating the parameters of match scores that have to be fused using GMM[2].

2) Subjective variation of number of mixtures depending on the modalities used.

3) Number of iterations required for Monte Carlo simulation convergence based on statistics obtained using GMM and also due to different modalities used.

2 Proposed Method

This section describes the proposed method for multimodal biometric score fusion. The fundamental idea behind the proposed method is that, we first estimate the statistics of match scores distribution using GMM. We then sample the match scores using Monte Carlo method. The observed values of statistic for these samples are then used to model the distribution. Finally, we use statistical hypothesis testing on sampled scores to decide whether user is genuine or imposter. Monte Carlo method has been used for centuries, but only in the past several decades has this technique gained the status of a full-fledged numerical method capable of addressing the most complex applications[10-11]. In many cases, we do not know accurately the sampling distribution for the statistic, or assumptions related to a particular nature of distribution are not fully satisfied. This is a typical case of biometric scores as the underlying distribution of genuine and imposter vscores are not

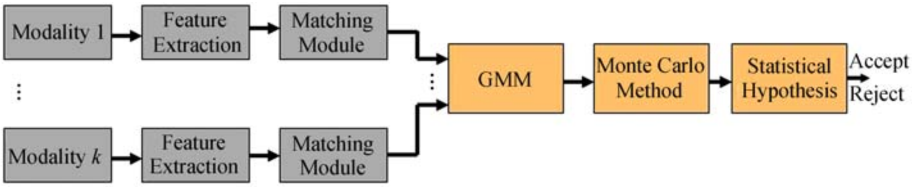


Fig.1. Block diagram of proposed method.

known accurately (even though for practical convenience the analysis is carried out by modeling sampled genuine and imposter scores to follow a Gaussian distribution). Prior to sampling we give maximum scope for the match scores to belong to arbitrary distribution and hence resort to a GMM process. Thus, rather than assuming the scores to follow a particular distribution and do fusion it will be better if we sample the score and perform fusion. In this context, we formulate the problem of person verification as a hypothesis testing problem with the following two hypothesis:

H_0 the person is Imposter
H_1 the person is Genuine.

Monte Carlo simulation can be used to evaluate the performance of a hypothesis in terms of the False Acceptance Rate (FAR) also called Type I error and False Reject Ratio (FRR) also called Type II error. We use upper tail (right tail) Monte Carlo hypothesis testing[11]. The probability of rejecting H_0 when H_0 is true is known as FAR and probability of not rejecting H_0 when it is false is known as FRR. Finally, Genuine Acceptance Rate (GAR) is found using $1 - FRR$. For fixed values of FAR, we estimate the GAR using Monte Carlo method. Before performing Monte Carlo simulation, it is necessary to estimate the mean and standard deviation of data under consideration. We employ Gaussian Mixture Model (GMM) to obtain mean and variance of scores under consideration.

Fig.1 shows the block diagram of the proposed method. The data obtained from each modality is used to extract the features. These extracted features are compared against the stored templates to generate match scores.

Let $\boldsymbol{X} = [X_1, X_2, \ldots, X_k]$ be the composite concatenated data of genuine and imposter scores obtained from k different matchers and each of these matchers have both genuine scores $\boldsymbol{X}_G$ and imposter scores $\boldsymbol{X}_I$. Let $\boldsymbol{X}_G$ denote the genuine scores obtained from "k" different matchers such that $\boldsymbol{X}_G = [X_{G1}, X_{G2}, \ldots, X_{Gk}]^{\mathrm{T}}$ and $\boldsymbol{X}_I$ be the imposter scores obtained from the same "k" different matchers such that $\boldsymbol{X}_I = [X_{I1}, X_{I2}, \ldots, X_{Ik}]^{\mathrm{T}}$. Based on the scores $\boldsymbol{X}_G$ and $\boldsymbol{X}_I$, we estimate the mean and covariance using GMM as follows.

2.1 Gaussian Mixture Model (GMM)

Density estimation of the match scores obtained from the different unimodal biometric system is challenging as they exhibit multimodality and there will be overlapping in tail region of genuine and imposter scores. Moreover, the presence of discrete scores in both genuine and imposter category makes the score level fusion a challenging task. In [12-13] to accurately estimate the density, non parametric technique like density histogram and kernel density estimation is employed. But these techniques require careful choice of histogram bandwidth or kernel bandwidth for accurate estimation of the match scores. GMM has been successfully employed in [14-15] to estimate the parameters of arbitrary distribution of the scores. Thus, GMM can handle a situation where a single parametric family is unable to provide a satisfactory model for local variation in observed data. The GMM can be written as[16-17]:

$$G(x) = \sum_{i=1}^{M} p_i b_i(x) \tag{1}$$

where

$$b_i(x) = \frac{1}{(2\pi)^{D/2}|\Sigma_i|^{1/2}} \exp\left\{-\frac{1}{2}(x-\mu_i)^{\mathrm{T}}\Sigma_i^{-1}(x-\mu_i)\right\}. \tag{2}$$

Here, μ_i, Σ_i represent the mean and covariance respectively of the i-th mixture. Given the scores of multimodal biometrics ($\boldsymbol{X} = [\boldsymbol{X}_G, \boldsymbol{X}_I]$) and number of mixtures, we estimate the statistics of genuine and imposter scores separately using GMM. Thus, the estimated mean for a given multimodal biometric scores with GMM will be in the form of μ_G and μ_I.

Here, μ_G indicates the composite mean of the genuine scores and μ_I indicates the composite mean of the imposter scores. We then estimate the covariance between genuine and imposter scores. Let "ρ" represent the estimated covariance between genuine and imposter scores, we then obtain the standard deviation as $\sqrt{\rho}$. Thus, the values obtained from GMM, such as mean of genuine scores, mean of imposter scores and standard deviation between genuine and imposter scores are used to sample the scores (from different matchers) using Monte Carlo method, which is described in the following section. In our experiments, we find the model order (number of mixtures) by cross validation. Given a dataset, we evaluate the performance over different numbers of mixture components. We then select the number of mixture components which gives the best performance. As shown in the Appendix, these parameters are subjective and it also depends on the type of modalities used and the class of algorithm used to evaluate. As such, we are trying to employ Monte Carlo methods to sample the scores in as much unbiased a manner as possible. Also such subjectivity is best handled by experimental approach as is done in [2, 18].

2.2 Monte Carlo Method to Estimate GAR

The following steps outline the procedure employed to estimate GAR.

Step 1. Let $\boldsymbol{X} = [\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3, \ldots, \boldsymbol{X}_k]$ denote the match scores obtained by k different matchers.

Step 2. Normalize the scores obtained from different matchers such that scores from different matchers lie in same range. In this work, we employed Min-Max normalization as it is known to be more effective[1].

Step 3. Using GMM we estimate the mean of genuine scores, the mean of imposter scores and standard deviation between genuine and imposter scores from different matchers.

Step 4. Determine a pseudo-population of interest where the null hypothesis is false.

Step 5. Generate random sample of size n (we experimentally fix value of $n = N_{\text{Gen}}$, i.e., the number of genuine scores) from this pseudo-population[1].

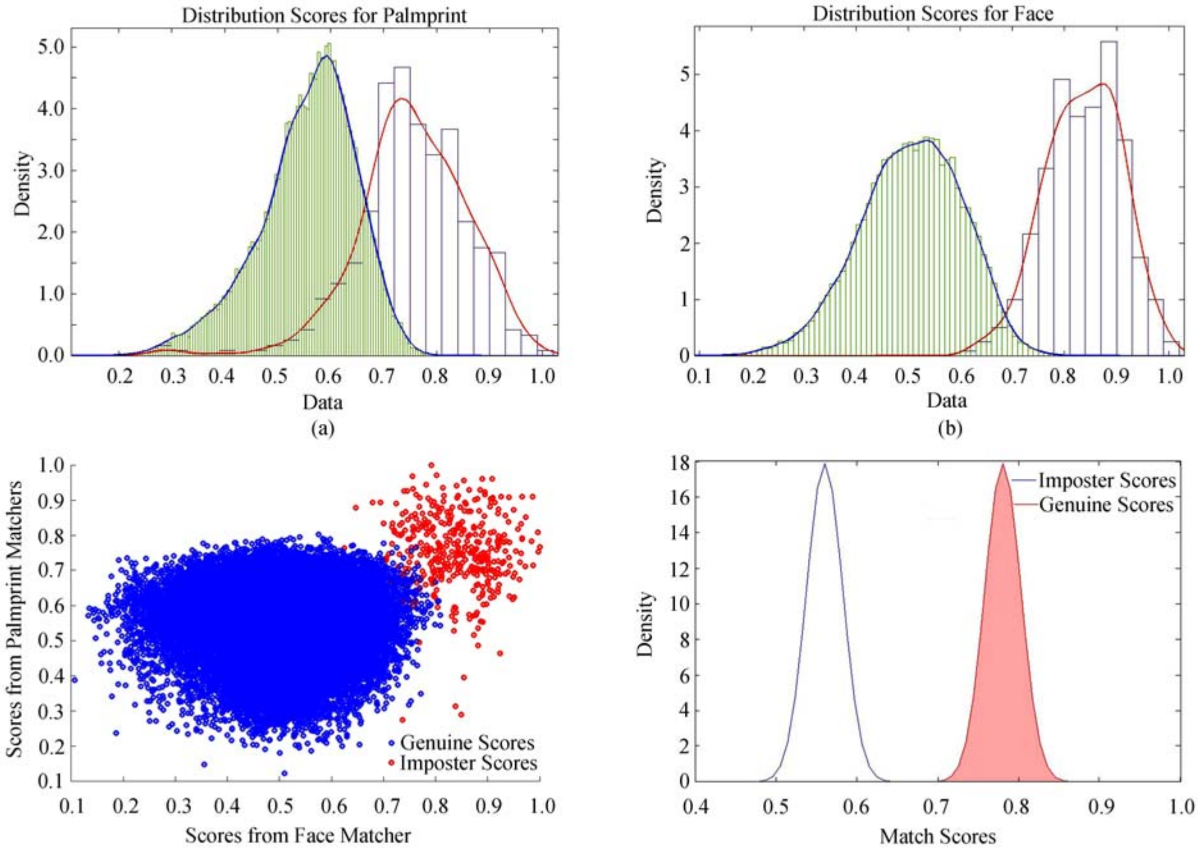
Step 6. For a given FAR get the critical value using Z-test[19]. We use Z-test as mean and standard deviation of the fused sample data is known from step 3.

Step 7. Perform hypothesis testing for the given FAR and corresponding critical value as obtained in step 6.

Step 8. Record the result (D_i) for this trail as:

$$D_i = \begin{cases} 1, & \text{person is genuine,} \\ 0, & \text{person is imposter.} \end{cases}$$

Step 9. Repeat steps 5~8 for N trials (in our work, we use $N = 25000$).



(a) (b) (c) (d)

Fig.2. Density estimation using Monte Carlo method. (a) Distribution of match scores for palmprint. (b) Distribution of match scores for face. (c) Scatter plot of genuine and imposter scores of palmprint and face. (d) Distribution of genuine and imposter scores of palmprint and face obtained using Monte Carlo method at $FAR = 0.1\%$.

Step 10. Then estimate $\widehat{FRR}$ using:

$$\widehat{FRR} = \frac{1}{N}\sum_{i=1}^{N} D_i. \tag{3}$$

Step 11. Obtain GAR using $GAR = 1 - \widehat{FRR}$.
There is no general theory that governs the number of trails in Monte Carlo simulation[20]. Usually, the match score distribution of individual biometrics have a long tail[1] and hence it is required to capture information contained in the tail of the distribution. As we are interested in capturing the information present in tail of score distribution, it is required to run Monte Carlo simulation for a large number of trails to ensure that there will be a good estimate of that portion. Thus in our analysis, we have chosen a value of $N = 25000$, i.e., the number of trails in Monte Carlo simulation. Appendix indicates not only the requirement of large number of trials but also the subjective variation (around the same order of magnitude) depending on the modalities used. Fig.2 shows the density estimation using proposed Monte Carlo method. The proposed method is illustrated by considering the example of fusing the match scores of individual biometrics such as palmprint and face. Fig.2(a) shows the match score distribution of palmprint alone while Fig.2(b) shows the distribution of match scores for face alone. Thus, it is observed from both Figs. 2(a) and 2(b) that estimation of these match scores are very difficult as it does not follow any particular distribution. The proposed Monte Carlo method will address this problem by sampling the distribution in such a manner that the sampled scores will follow a Gaussian distribution. We then employ statistical hypothesis to make a decision about accept/reject. Fig.2(c) shows the scatter plot of genuine and imposter scores of face and palmprint matchers. Fig.2(d) shows the resulting distribution of genuine and imposter scores obtained by fusing the match scores of face and palmprint using proposed method. It is observed that the sampled fused scores follows a Gaussian distribution and it is also evident from Fig.2(d) that the proposed method has overcome two important difficulties in handling biometric score distribution such as heavy tail and multiple modes.

3 Experimental Setup

This section describes the experimental setup to evaluate the proposed Monte Carlo method of match score fusion. The need for going in for an experimental approach is quite evident from [1-2] and further justified in Appendix of this paper. The experiments are carried out on five different multimodal biometric databases, out of which, one is NIST database and other four are built as multimodal biometric databases. In building the multimodal biometric databases, the biometric trait of a user from a database is combined with another biometric trait of yet another user as indicated in [1]. The five multimodal biometric databases used in our experiments are summarized in Table 1 and brief description of these five different multimodal databases are described below.

3.1 NIST-Multimodal

We use NIST Biometric Score Set, Release 1 (BSSR 1)[21] to validate the proposed method of score fusion. BSSR set 1 holds the multimodal biometric match scores of face and fingerprint data from 517 users. These scores are collected from two anonymous 2002 face matchers (C and G), and one fingerprint matcher (V). Thus, BSSR set 1 contains match scores obtained from: right index fingerprints scored by matcher (V), left index fingerprints scored by matcher (V), frontal face images scored by matcher C and frontal face image scored by matcher G.

3.2 Build Multimodal Biometric Database-1 (BMBD-1)

This Database is built using face, palmprint and speech. For face modality, we employ AR database[22]. This consists of 126 users face images, including different facial expression, occlusion and lighting condition.

Table 1. Summary of Multimodal Biometric Databases

Database	Biometric Traits	No. Matchers	No. Users
NIST-Multimodal	FingerPrint (Two Fingers) Face (Two matchers)	4	517
Build Multimodal Biometric Database-1	Face, Speech, Palmprint	3	100
Build Multimodal Biometric Database-2	Face, Palmprint	2	100
Build Multimodal Biometric Database-3	Face, Speech	2	100
Build Multimodal Biometric Database-4	Palmprint, Speech	2	100

The pictures of most persons were taken in the two sessions. Since, we are using AR database to evaluate the performance of algorithms for facial expression and time variation, we select full facial image corresponding to facial expression in both sessions by discarding images corresponding to occlusion and illumination. For our experiments, we select 100 users and each user has 10 facial images such that 5 images are considered from first session and next 5 images are from the second session. Out of 10 facial images, 6 images are used for training (3 from the first session and 3 from the second session) and 4 images (2 from the first session and 2 from the second session) are used for testing. For palmprint modality, we select a subset of 100 users from PolyU database[23], each of these users has 10 samples such that 5 samples are taken from the first session and next 5 samples are taken from the second session. Then, out of 10 samples, 6 samples (3 samples from the first session and 3 samples from the second session) are used for training, 4 samples, (2 samples from the first session and 2 samples from the second session) are used for testing. For speech modality, the testing corpus of 100 users from TIMIT database[24] is used. Each user has 10 samples out of which 6 samples are used for training and 4 samples are used for testing. To obtain the match scores, each of these modalities are analyzed using different feature extraction algorithms that are outlined in the following subsections.

3.2.1 *Face Verification*

The Log Gabor transform[25] is used to extract the texture information of face image. On a linear frequency scale, the transfer function of Log Gabor behavior has the form[25]:

$$G(\omega)=\exp\left\{\frac{-\log(\omega/\omega_o)^2}{2\times\log(k/\omega_o)^2}\right\} \tag{4}$$

where ω_o is the filters center frequency. To obtain a constant shape filter, the ratio k/ω_o must be held constant for varying ω_o.

In our experiments, all the face images are of size 60×60 pixels. The Log Gabor filter used in our work has 4 different scales and 8 different orientations. Thus, each facial image is filtered using 8×4 different Log Gabor filters that results in 32 different filtered images. So, each facial image is represented using 32 different images. This will result in a dimension of 240×480 for each facial image. Finally, Nearest Neighbor Classifier (NNC)[17] is used to obtain the match scores.

3.2.2 *Palmprint Verification*

In our work, the eigenpalm approach is used[26]. The eigenpalm based palmprint verification is divided into two stages: 1) training stage, 2) testing stage. In the training stage, a set of palmprint images that best describe the distribution of the training palmprint images in lower dimensional subspace (eigenspace) is computed. Using PCA, we project the training palmprint images onto the eigenspace[25]. In testing stage, a test palmprint image is projected onto the same eigenspace and the match score between input palmprint image and test palmprint image is computed using NNC. In our experiments, the Region Of Interest (ROI) is extracted for all the palmprint images and are of size 60×60 pixels. Then, PCA is applied on these palmprint images[26].

3.2.3 *Speaker Verification*

In our work, text independent speaker verification is used. The text independent speaker verification uses sphericity based measure[27-28]. The speech signal is analyzed on a frame by frame basis, with a typical frame length of 25ms and frame advance by 15ms. For each frame, 12-dimension Linear Prediction Cepstral Coefficients (LPCC) is extracted. Each client is modeled by the covariance matrix $\boldsymbol{X}$ of the feature vectors of client training data $\{\boldsymbol{X}_1,\boldsymbol{X}_2,\ldots,\boldsymbol{X}_N\}$, where

$$\hat{\boldsymbol{X}}=\frac{1}{N-1}\sum_{j=1}^{N}\boldsymbol{X}_j \tag{5}$$

and then,

$$\boldsymbol{X}=\frac{1}{N-1}\sum_{j=1}^{N}(\boldsymbol{X}_j-\hat{\boldsymbol{X}})(\boldsymbol{X}_j-\hat{\boldsymbol{X}})^{\mathrm{T}}. \tag{6}$$

During the test session, the covariance matrix $\boldsymbol{Y}$ is computed over test speech data of a person under test. The arithmetic-harmonic sphericity measure $S(\boldsymbol{X},\boldsymbol{Y})$[28] is used to make the final decision about the person under test. The similarity measure between the client and the person under test using arithmetic-harmonic sphericity measure is given as:

$$S(\boldsymbol{X},\boldsymbol{Y})=\log\frac{tr(\boldsymbol{Y}\boldsymbol{X}^{-1})tr(\boldsymbol{X}\boldsymbol{Y}^{-1})}{m^2} \tag{7}$$

where m is the dimension of the feature vector and $tr(y)$ is the trace of the y. Specific algorithms for face, palmprint and speech described above are appropriately used on compatible databases. The resulting match scores are used to evaluate the performance.

3.3 Build Multimodal Biometric Database-2 (BMBD-2)

This database is built using face and palmprint. Face and palmprint samples are obtained as described in

Subsection 3.2. Thus, this database has 100 users such that each users has 10 samples of face and 10 samples of palmprint. The algorithm for palmprint and face described in Subsection 3.2 are used to obtain match scores.

3.4 Build Multimodal Biometric Database-3 (BMBD-3)

This database is built using face and speech. Face samples are obtained as described in Subsection 3.2.1 and speech samples are obtained as described in Subsection 3.2.3. Thus, this database has 100 users such that each user has 10 samples of face and 10 samples of speech. The algorithm for speech and face described in Subsection 3.2 is used to obtain match scores.

3.5 Build Multimodal Biometric Database-4 (BMBD-4)

This database is built using palmprint and speech samples. Here also, the algorithm for speech and palmprint described in Subsection 3.2 is used to obtain match scores. This database is of size 100 users such that each user has 10 samples of palmprint and 10 samples of speech.

4 Results and Discussion

This section describes the results obtained using proposed Monte Carlo method of fusion on five different build multimodal databases. In conducting the experiments on each of these databases, half of genuine and half of imposter scores are randomly selected to estimate the statistics of scores using GMM. This partition is repeated for "m" times ($m = 30$) and reported ROC curves correspond to mean GAR values over "m" trails at different FAR values. On each of these databases, the results of proposed fusion method are compared with results from two well known fusion methodology such as weighted sum rule[1] and LR method[2]. In performing weighted sum rule, the weights are assigned to each of the biometric based on their performance as described in [29].

4.1 Results on NIST Database

Fig.3 shows the performance of the proposed method along with individual biometrics on NIST multimodal biometric database. As proposed method requires setting of null and alternate hypothesis, following procedure is adopted: the null hypothesis (H0) is set by computing the composite mean of imposter scores obtained from individual biometrics and alternate hypothesis (H1) may be any value greater than null hypothesis. To effectively analyze the performance of the proposed scheme, we have to choose one optimum value for alternate hypothesis. This optimum value for alternate hypothesis is obtained by finding the composite mean of genuine scores of individual biometrics. This value is then varied between ±0.1 (in steps of 0.01) to effectively analyze the performance of proposed method and then we average the results obtained across all these hypothesis. This procedure is repeated for "m" number of trails and mean of the result is shown in Fig.3. The proposed method shows the best result of 99.82% at $FAR = 0.01\%$ as compared with best individual (unimodal) biometric trait, i.e., right index finger. In this experiment, we do not rerun the results of comparative methods such as weighted sum rule and LR method as these results are available in [2]. Table 2 summarizes the comparative results of our approach with other methods.

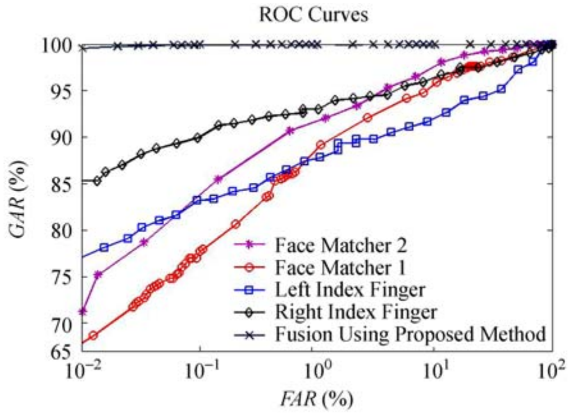


Fig.3. Performance of proposed method on NIST database.

4.2 Results on Build Multimodal Biometric Database-1

Fig.4 shows the performance of the proposed method on Build Multimodal Biometric Database-1. Here also it is observed that the proposed method outperforms all other methods with $GAR = 97.7\%$ at $FAR = 0.01\%$.

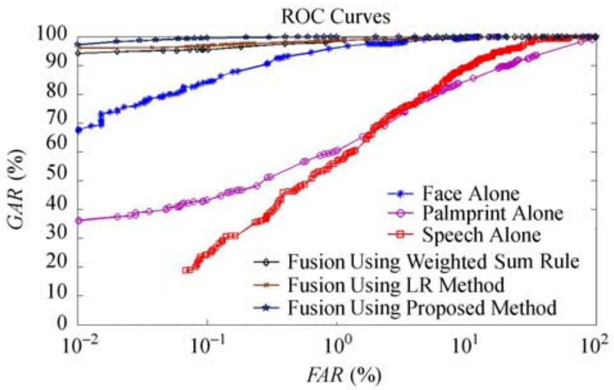


Fig.4. Performance of proposed method on Build Multimodal Biometric Database-1.

The second best performance is noted for LR method with $GAR = 96.4\%$ at $FAR = 0.01\%$. Similarly, as in previous database, here also the null hypothesis is set by computing composite mean of imposter scores, where as optimum alternative hypothesis is set by computing composite mean of genuine scores. Finally, this value of optimum alternative hypothesis is varied between ±0.1 (in steps of 0.01) and we then consider the average of results obtained across all these hypothesis. This procedure is repeated for "m" number of trails and mean of the result is shown in Fig.4.

4.3 Results on Build Multimodal Biometric Database-2

Fig.5 shows the performance of the proposed method on Build Multimodal Biometric Database-2. Here the proposed method shows an improvement of 3.80% in GAR at $FAR = 0.01\%$ as compared with the second best method, i.e., LR Method. To set the values for null and optimum alternative hypothesis, we follow procedure similar to that as in Subsection 4.2. Fig.5 shows the ROC curves corresponding to mean GAR values obtained over "m" trails at different FAR values.

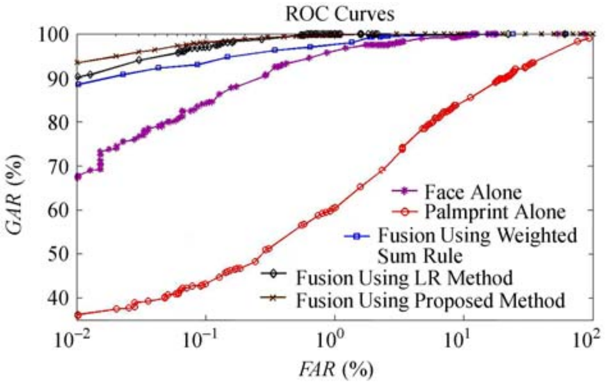


Fig.5. Performance of proposed method on Build Multimodal Biometric Database-2.

4.4 Results on Build Multimodal Biometric Database-3

Fig.6 shows the performance of proposed method on Build Multimodal Biometric Database-3 composed of face and speech. It is observed from Fig.6 that the proposed method outperforms the remaining methods with $GAR = 96.4\%$ at $FAR = 0.01\%$. To set the values for null and optimum alternative hypothesis, we follow the procrdure similar to that as indicated in Subsection 4.2. Fig.6 shows the ROC curves corresponding to mean GAR values obtained over "m" trails at different FAR values.

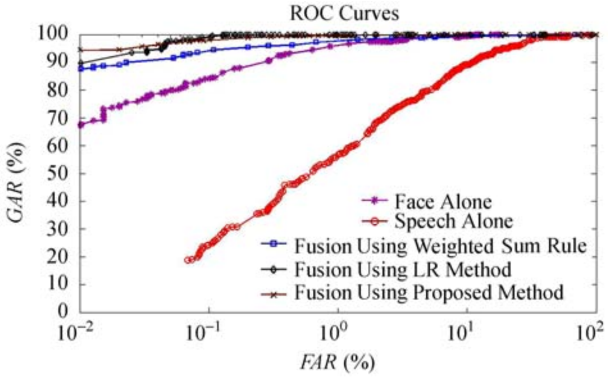


Fig.6. Performance of proposed method on Build Multimodal Biometric Database-3.

4.5 Results on Build Multimodal Biometric Database-4

Fig.7 shows the performance of proposed method on Build Multimodal Biometric Database-4. It is seen that the proposed method shows an improvement of GAR by 28.38% at $FAR = 0.01\%$ as compared with second best method, i.e., LR method. Like earlier cases, here also the null hypothesis in each experiment is set by computing composite mean of imposter scores and optimum alternate hypothesis is set by computing composite mean of genuine score. Then, this optimum hypothesis value is varied between ±0.1 (in steps of 0.01) and we then average the results obtained across all these hypothesis. This procedure is repeated for "m" number of trails and average of the result is shown in Fig.7.

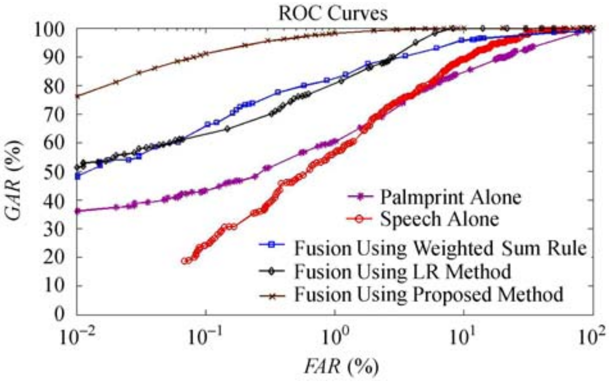


Fig.7. Performance of proposed method on Build Multimodal Biometric Database-4.

As observed from Figs. 3~7, the proposed method outperforms all other methods on different databases with varying number of matchers, users and also with different biometric traits. Table 2 summarizes the performance of the proposed method, where performance is indicated by GAR value obtained at $FAR = 0.01\%$ and also the standard deviation obtained from "m" number

Table 2. Comparison with the Proposed Method

Database	Mean GAR (%) at $FAR = 0.01\%$ and Std. Deviation (%)		
	Proposed Method (GMM-MC)	Likelihood Ratio Method	Weighted Sum Rule
NIST-Multimodal	99.87 ± 0.06	99.1[2]	98.0[2]
BMBD-1	97.74 ± 0.04	96.43 ± 0.12	95.39 ± 0.13
BMBD-2	93.95 ± 0.02	90.15 ± 0.21	89.67 ± 0.15
BMBD-3	96.46 ± 0.02	90.06 ± 0.31	89.85 ± 0.12
BMBD-4	78.47 ± 0.04	50.09 ± 0.18	49.63 ± 0.19

of trails. Thus, from Table 2 it is clear that the proposed method shows very consistent improvement across the 30-cross validation trails and it shows a very small value of deviation (less that 0.1%). It is also observed that the improvement achieved by LR method over weighted SUM rule is very less, while the proposed method shows signification improvement over these two schemes on five different databases.

5 Conclusion

In this paper, we propose a framework for biometric score fusion using GMM and Monte Carlo sampling method. Unlike existing multimodal biometric approaches, the proposed approach utilizes the statistics of scores and generates large number of samples, then, we set statistical hypothesis that result in performance better than existing techniques. GMM is used to estimate the statistics of the multimodal biometric scores under consideration which in turn is used by Monte Carlo method. The efficacy of the proposed method is evaluated on five build multimodal biometric databases. Based on the extensive experiments we can conclude that, the proposed method of fusion achieves consistently higher verification rates as compared with most popular and state-of-the-art fusion schemes such as weighted sum rule[1] and likelihood ratio method[2] on all the five different multimodal biometric databases.

Acknowledgement The authors thank the three anonymous reviewers of an earlier version of this paper. Their comments and suggestions greatly helped us to improve the presentation and explanation.

References

[2] Nandakumar K, Chen Y, Dass S C, Jain A K. Likelihood ratio-based biometric score fusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2008, 30(2): 342-347.

[1] Ross A, Nandakumar K, Jain A K. Handbook of Multibiometrics. Springer-Verlag, 2006.

[3] Kittler J, Hatef M, Duin R P W, Matas J. On combining classifiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1998, 20(3): 226-239.

[4] Jain A K, Hong L, Kulkarni Y. A multimodal biometric system using finger-print, face, and speech. In *Proc. International Conference on Audio and Video Based Biometric Person Authentication*, Washington DC., USA, 1999, pp.182-187.

[5] Snelick R, Uludag, U, Mink A, Indovina M, Jain A K. Large-scale evaluation of multimodal biometric authentication using state-of-the-art systems. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2005, 27(3): 450-455.

[6] Ross A, Jain A K. Information fusion in biometrics. *Pattern Recognition Letters*, 2003, 24(13): 2115-2125.

[7] Chu R, Liao S, Han Y, Sun Z, Li S Z, Tan T. Fusion of face and palmprint for personal identification based on ordinal features. In *Proc. IEEE Conference on Computer vision and Pattern Recognition*, Minneapolis, USA, June 18-23, 2007, pp.1-2.

[8] Dass S C, Nandakumar K, Jain A K. A principled approach to score level fusion in multimodal biometric systems. In *Proc. Audio and Video based Biometric Person Authentication* (*AVBPA-2005*), New York, USA, July 20-22, 2005, pp.1049-1055.

[9] Fierrez-Aguilar J, Ortega-Garcia J, Gonzalez-Rodriguez J. Fusion strategies in multimodal biometric verification. In *Proc. ICME 2003*, Baltimore, USA, July 6-9, 2003, pp.5-8.

[10] Sobel I M. A Primer for the Monte Carlo Methods. CRC Press, 2000.

[11] Kalos M H, Whitlock P A. Monte Carlo Methods. Wiley, 1986.

[12] Li J Q, Barron A. Mixture Density Estimation. Advances in Neural Information Pressing System, 12, MIT Press, 1999.

[13] Rakhlin A, Penchanko D, Mukerjee S. Risk bounds for mixture density estimation. *ESAIM: Probability and Statistics*, 2005, 9: 220-229.

[14] Ulery B, Hicklin A R, Watson C, Fellener W. Studies of biometric fusion. Technical Report IR 7346, NIST, 2006.

[15] Figueiredo M, Jain A K. Unsupervised learning of finite mixtures modesl. *PAMI*, 2002, 24(3): 381-396.

[16] McLachlan G J, Peel D. Finite Mixture Models. Wiley Series in Probability and Statistics, 2000.

[17] Bishop C. Pattern Recognition and Machine Learning. Springer, 2006.

[18] Raghavendra R, Rao A, Hemantha Kumar G. Subjective performance of texture based algorithm for face vereification: The role of databases. In *Proc. the 6th Indian Conference on Computer Vision, Graphics & Image Processing* (*ICVGIP 2008*), Hyderabad, India, Jan. 11-13, 2000, pp.421-428.

[19] Ross S M. Introduction to Probability and Statistics for Engineers and Scientists. Third Edition, Elsvier Academic Press, 2004.

[20] Mooney C Z. Monte Carlo Simulation. Sage Publications, 1997.

[21] NIST biometric score sets-release 1. Nat. Inst. Standards and Technology, 2004, http:/www.itl.nist.gov/iad/894.03/biometricscores.

[22] The AR Face Database. www.rvl1.ecn.purdue.edu.aleix face DB.html/.

[23] polyU palmprint database. www.comp.polyu.edu.hk/biometrics.

[24] TIMIT speech database. www.ldc.upenn.edu.

[25] Xiao Z, Guo C, Yu M, Li Q. Research on log Gabor wavelet and its application in image edge detection. In *Proc. the 6th International Conference on Signal Processing* (*ICSP-2002*), Beijing, China, Aug. 26-30, 2002, pp.592-595.

[26] Zhang D. Palmprint Authentication. Springer-Verlag, 2004.

[27] Ben-Yacoub, S, Abdeljaoued Y, Mayoraz E. Fusion of face and speech data for person identity verification. *IEEE Transactions on Neural Networks*, 1999, 10: 1065-1074.

[28] Bimbot F, Magrin I, Mathan L. Second order statistical measure for text independent speaker identification. *Speech Communication*, 1995, 17(1/2): 177-192.

[29] Raghavendra R, Rao A, G Hemantha Kumar. Qualitative weight assignment for multimodal biometric fusion. In *Proc. the 7th International Conference on Advances in Pattern Recognition* (*ICAPR 2009*), Kolkata, India, Feb. 4-6, 2009, pp.193-196.

R Raghavendra received the B.E and M.Tech degrees in electronics and communication engineering from University of Mysore and Visvesvaraya Technological University respectively. Since 2007, he has been a candidate for Ph.D. degree in computer science and technology in University of Mysore. His current research interest includes DSP, pattern recognition, feature selection, optimization techniques for biometrics and finite mixture models. He has authored more than 20 research publications in all areas of his interest in reputed conferences and journals.

Rao Ashok received his B.E., M.E. and Ph.D. degrees in 1982, 1985, 1991, all in EE from University of Mysore, IISc Bangalore and IIT Bombay respectively. His current area of research includes biometrics, DSP and image processing, bio-medical signal processing, applied linear algebra, social computing, renewable energy and engineering education. From 1999~2005 he was head of Network Project, CEDT, IISc. From August 2009, he is professor, Dept. E & C and head, Education and Research at CIT Gubbi, Tumkur. He has authored over 75 research publications in all areas of his interest.

G Kumar Hemantha received B.Sc., M.Sc and Ph.D. degrees from University of Mysore. He is working as a professor in the Department of Studies in Computer Science, University of Mysore, Mysore. He has published more than 200 papers in journals, edited books and refereed conferences. His current research interest includes numerical techniques, digital image processing, pattern recognition and multimodal biometrics.

Appendix. Description of the Proposed Method

In this Appendix, we provide detailed analysis of the proposed method for multimodal biometric fusion by considering BMBD 4 database (palmprint and speech) as an example. The proposed method can be considered in the following three steps:

Step 1.

Given the palmprint and speech modalities we first obtain the genuine and imposter scores. We then employ Gaussian Mixture Model (GMM) to estimate the underlaying statistics such as mean and covariance of genuine and imposter scores. Table A1 indicates the statistics of the genuine and imposter scores obtained for different number of mixtures. This exhibits typical issue of overfitting or underfitting of data as can be seen from Table A1. For this choice of modality and databases, GMM modeling using 1, 2, 3 mixture components appear a case of underfitting and GMM mixture 5, 6 appear to be overfitting. This endorses the need to experimentally fit the best order model. This obviously depends on database, modality used, etc.[18] Choosing the number of mixtures to be one, amounts to modeling the process by a single Gaussian.

Table A1. Statistics of Genuine and Imposter Scores Obtained for Different Mixtures of GMM

Number of Mixtures in GMM	1	2	3	4	5	6
Mean of Genuine Scores	0.72	0.74	0.75	0.75	0.73	0.73
Mean of Imposter Scores	0.68	0.64	0.60	0.58	0.63	0.68
Covariance Between Genuine and Imposter Scores	0.037	0.032	0.028	0.025	0.042	0.036

Step 2.

We then use these statistics estimated by GMM to sample the fused scores using Monte Carlo method. Thus, the sampled fused scores will follow the Gaussian distribution with the same statistics as that of the original scores. Fig.A1 shows the sampled fused scores obtained using Monte Carlo method for different number of mixtures.

Step 3.

In this step, we perform the statistical hypothesis testing to decide whether claimed user can be accepted or rejected. Fig.A2 shows the ROC curves obtained on fused scores with different number of mixtures. Thus, it is observed from Fig.A2 that the best result is noted when the number of Gaussian mixtures is 4 with *GAR* = 77.77% at *FAR* = 0.01%. This is also supported by way of Fig.A2, which shows for 4 mixtures the "best"

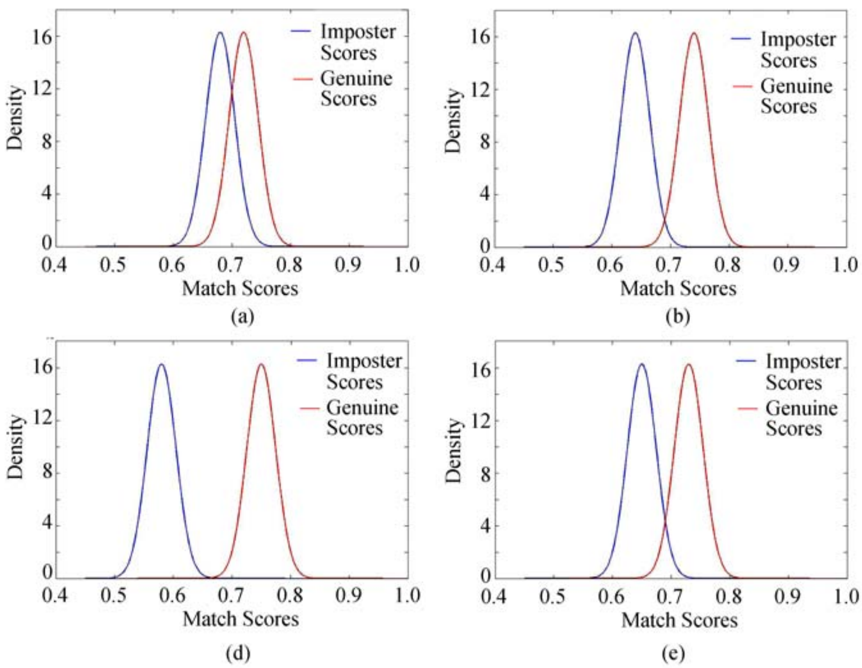


Fig.A1. Density estimation using Monte Carlo method at $FAR = 0.1\%$. (a) No. Mixture = 1. (b) No. Mixture = 2. (c) No. Mixture = 3. (d) No. Mixture = 4. (e) No. Mixture = 5. (f) No. Mixture = 6. For $N = 25000$, $n = N_{\text{Gen}}$.

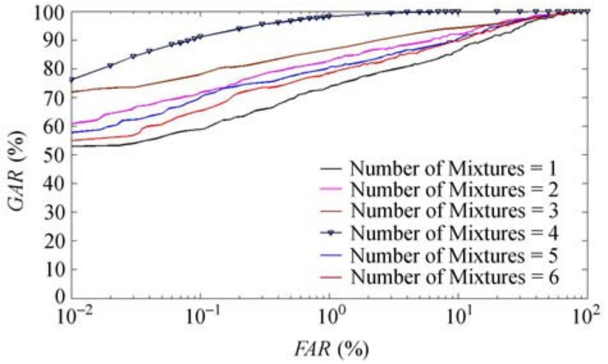


Fig.A2. Performance variation for different number of mixtures.

separation of imposter and genuine sampled score distribution. This further corroborated in Table A1 where, for this case, we have maximum mean of genuine scores, minimum mean of imposter scores and minimum covariance between genuine and imposter scores.

To Select the Value of N

Fig.A3 shows the variation of performance with respect to number of iteration used on BMBD-4 with 4 mixtures. It is observed that the performance of the proposed algorithm increases with number of iteration till it reaches some optimum value, for example, here it is 25000 then afterwards the results are stabilized with respect to N i.e., number of iteration. So based on this we select 25000. Of course this can vary due to different choice of modality or on different database.

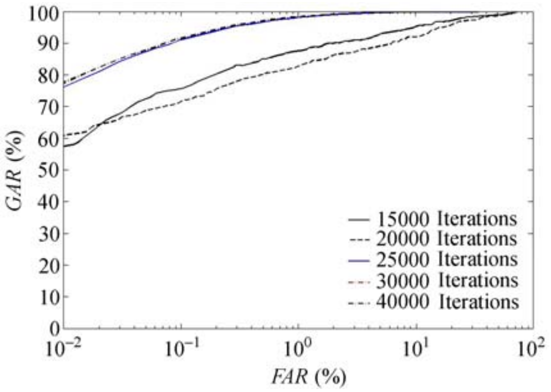
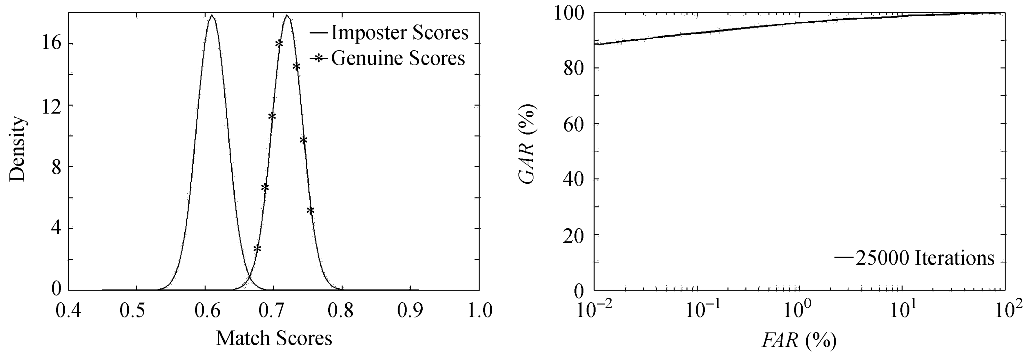


Fig.A3. Variation of performance vs. value of N.

Results on Other Databases

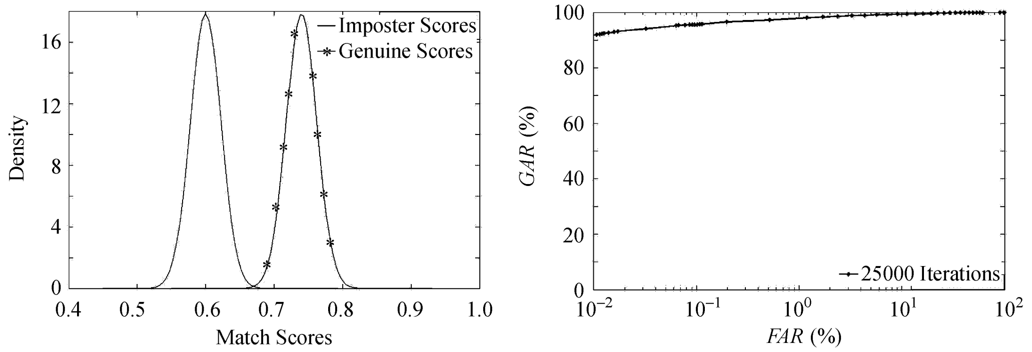
In order to check the universality of the parameters used in the proposed method, we conducted additional experiments and results are reported in this Appendix. In this experiment, we use the best parameters of proposed method that are obtained on BMBD-4 such as number of mixture = 4 and number of iteration = 25000.

Fig.A4(a) shows the sampled scores obtained from obtained on BMBD-2 (face and palmprint) and Fig.A5(b) shows the performance of the proposed method with 4 Gaussian mixtures and 25000 iterations. As compared to the results reported in Fig.5 and Table 2, results here indicate degraded performance with the present parameters. Thus, from the above analysis it is observed that the parameters of the proposed method has to be set experimentally and these parameters varies according to nature of the database. Similar results can also be noted on BMBD-3 database (face and speech) whose results are indicated in Figs. A5(a) and A5(b). Such subjective results on multimodal biometric databases are well known. The work in [18] shows in detail significant subjectivity of performance on various databases and various class of algorithms and methods employed. Our work here reinforces this further and also makes a very strong case for evaluation of algorithms and schemes based on extensive experiments.



(a) (b)

Fig.A4. Density estimation using Monte Carlo method at $FAR = 0.1\%$. (a) No. Mixture = 4. (b) ROC Curve for 25000 iterations on BMBD-2 database.



(a) (b)

Fig.A5. Density estimation using Monte Carlo method at $FAR = 0.1\%$. (a) No. Mixture = 4. (b) ROC Curve for 25000 iterations on BMBD-3 database.